

Briefing paper on Gates AR data

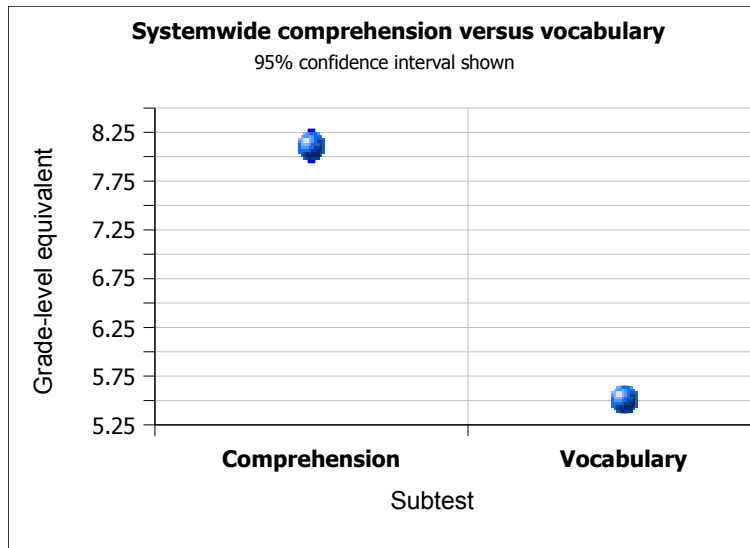
This briefing paper is a look at the Gates AR data gathered across all six sites at the College of Micronesia-FSM.

A number of statistically cautionary notes must be given up front. The test used, the Gates AR is not designed to determine grade level equivalents below the seventh grade. Mean grade level equivalents below seventh grade are simply beyond the design range of the instrument. Courses in this range would be better measured by other instruments.

A second note is that Gates AR notes that the grade level equivalent scores do not behave in a manner that ensure the results are standardized and are parametric. Technically, this means that these scores ought not be looked at with tools such as confidence intervals based on a parametric curves such as the student's t-distribution. The complication is that the alternate standardized, parametric scores have no particular meaning to an audience of college faculty and administrators. Grade level is the one measure that a general audience can connect with and understand. This also allows some crude "back-of-the-envelope" comparisons to measures of text grade level. Thus, despite the warnings, this document explores the grade level data as if it were amenable to confidence level statistics. The confidence intervals still provide some guidance as to the internal variation in the underlying data, and reflects the uncertainty that comes with small sample sizes.

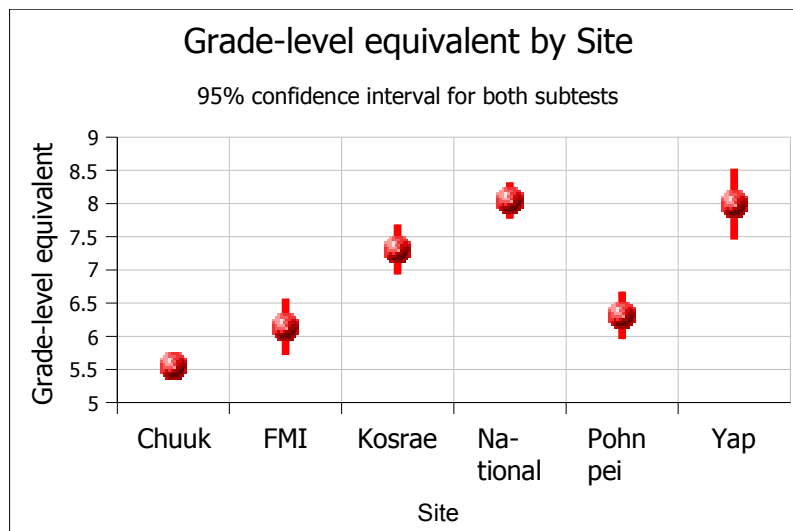
Comprehension and Vocabulary: System-wide, all courses

System-wide, for all courses, the mean comprehension grade-level equivalent exceeds the mean vocabulary grade-level equivalent by three grades. Students are stronger in comprehension than in vocabulary.

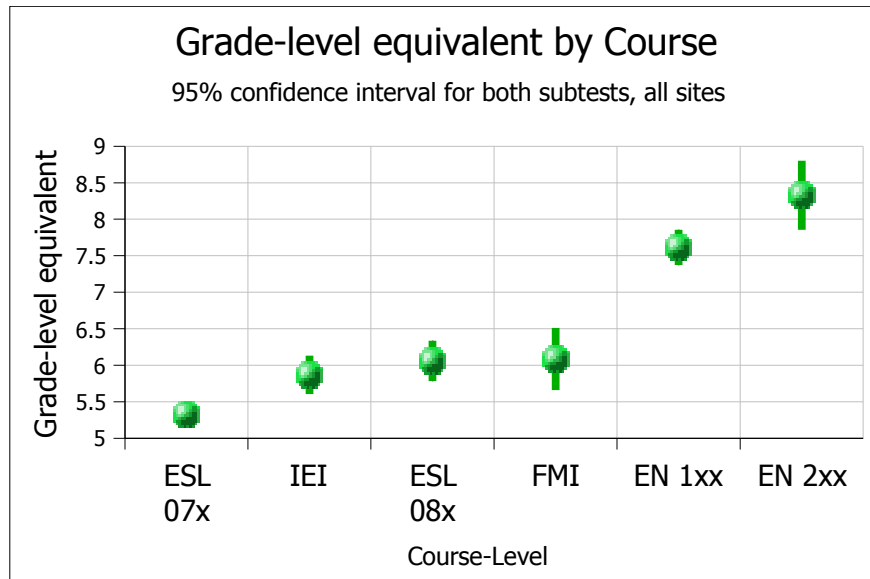


Overall mean grade-level equivalents by site for all courses

The mean grade-level equivalent for all students across all courses at a site varies by site. All sites face students with mean grade-level equivalents ranging from fifth to eighth grade.



Overall mean grade-level equivalents by course for all sites



The mean grade-level equivalent improves with course level. IEI (also known as ESL 086) and ESL 088/ESL 089 courses have roughly equivalent mean grade-levels. This suggests the courses are potentially redundant. All of the mean grade-level equivalents are well below that of an on-level 12th grade high school graduate, no mean exceeds a middle school level.

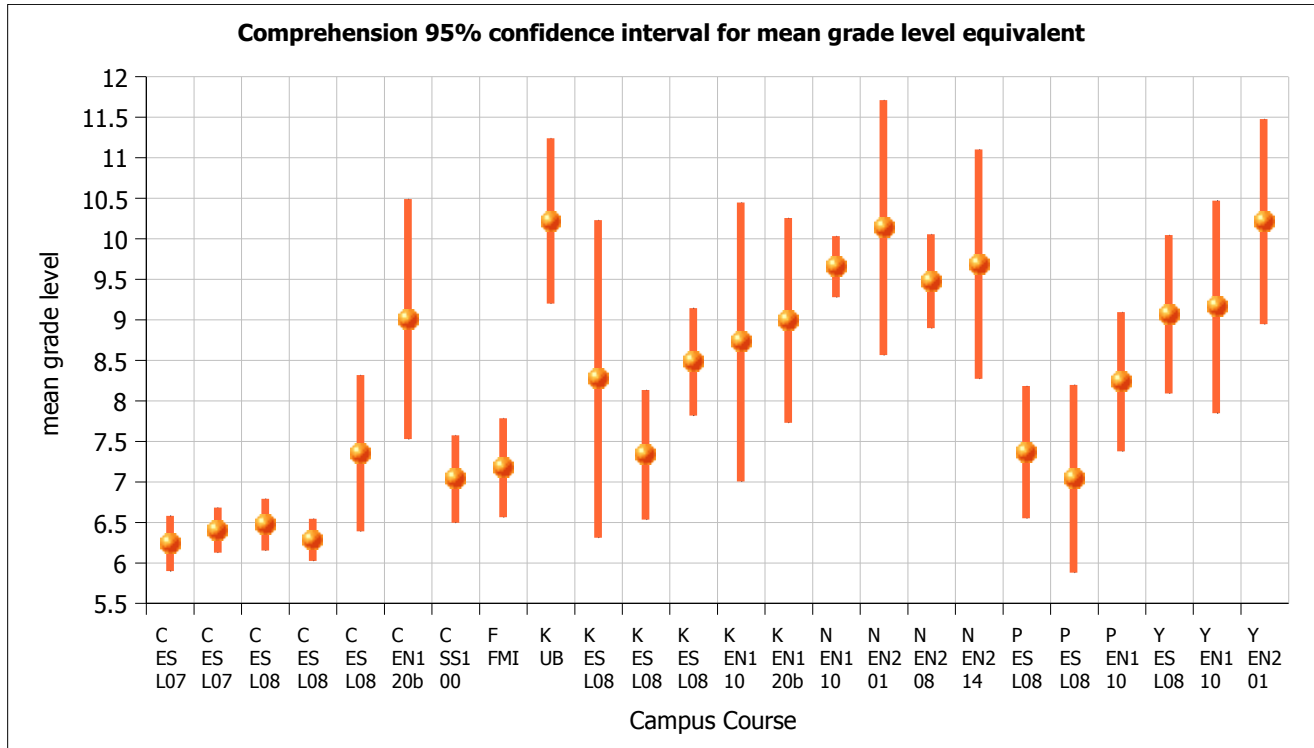
Comprehension: Details by campus and course

The following table contains statistics related to the mean comprehension score for various courses at the six sites of the college. The course appellations are those provided in the original spreadsheets, not all of them are familiar to this author.

Course	Campus	Sample size n	Standard deviation	Mean (average)
ESL070	Chuuk	41	1.08	6.24
ESL071	Chuuk	66	1.14	6.40
ESL086	Chuuk	75	1.38	6.47
ESL088	Chuuk	13	0.43	6.28
ESL089	Chuuk	17	1.88	7.35
EN120b	Chuuk	11	2.21	9.01
SS100	Chuuk	14	0.93	7.04
FMI	FMI	31	1.66	7.17
UB	Kosrae	20	2.18	10.22
ESL086	Kosrae	7	2.12	8.27
ESL088	Kosrae	14	1.39	7.34
ESL089	Kosrae	17	1.29	8.48
EN110	Kosrae	11	2.56	8.73
EN120b	Kosrae	14	2.19	8.99
EN110	National	104	1.94	9.66
EN201	National	13	2.6	10.14
EN208	National	44	1.9	9.48
EN214	National	15	2.56	9.69
ESL088	Pohnpei	18	1.64	7.37
ESL089	Pohnpei	8	1.39	7.04
EN110	Pohnpei	11	1.28	8.24
ESL086	Yap	19	2.03	9.07
EN110	Yap	15	2.37	9.16
EN110/201	Yap	3	1.4	10.5
EN201	Yap	16	2.37	10.21

Table 1: Comprehension grade-level equivalents

The comprehension data above is rendered as a chart on the next page.



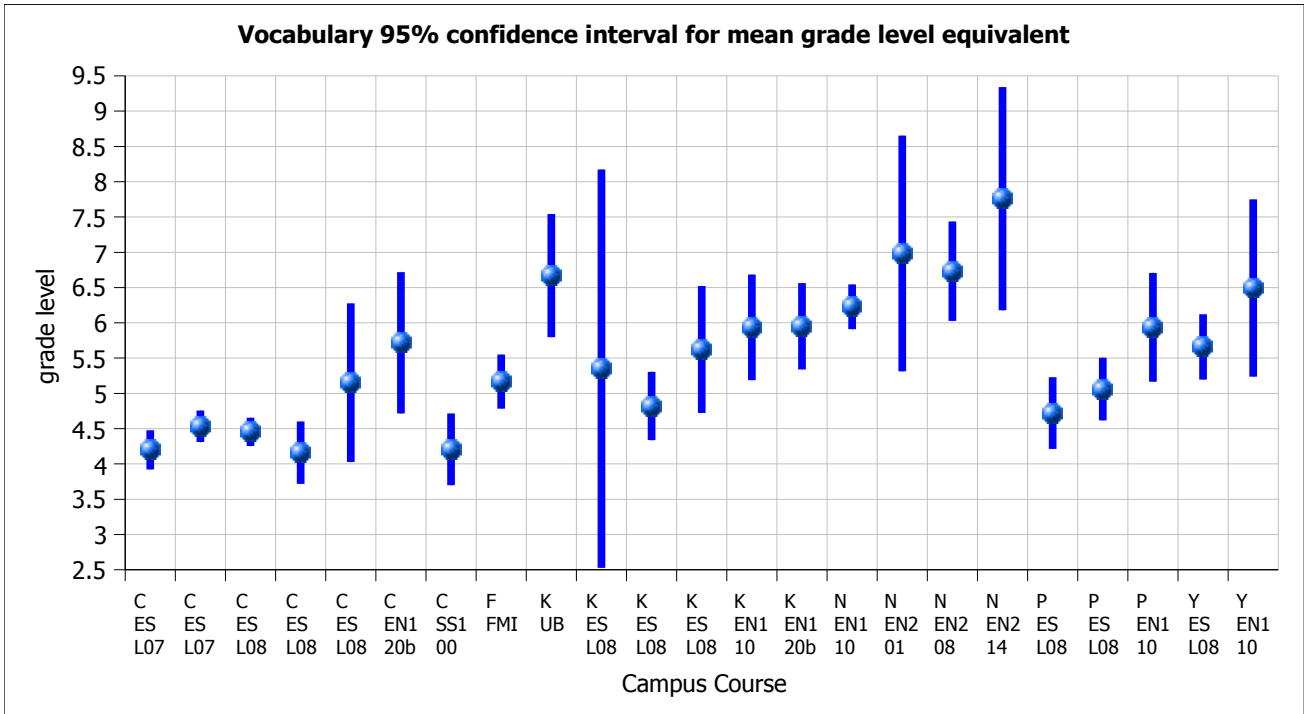
The chart makes more visually apparent that students in Chuuk start at a lower grade equivalent level than any other group of students at the other sites. This may in part be a sampling issue: no other campus tested their 070 level students. Of the three sites reporting performance of students in ESL 088 Chuuk site has the lowest grade level equivalent at 6.28 versus 7.34 for Kosrae and 7.37 for Pohnpei site. The Pohnpei site data has a large uncertainty due to a small sample size, which is apparent in the statistically insignificant drop in the mean for Pohnpei site from ESL 088 to ESL 089.

Vocabulary: Details by campus and course

The vocabulary means are lower than the comprehension means.

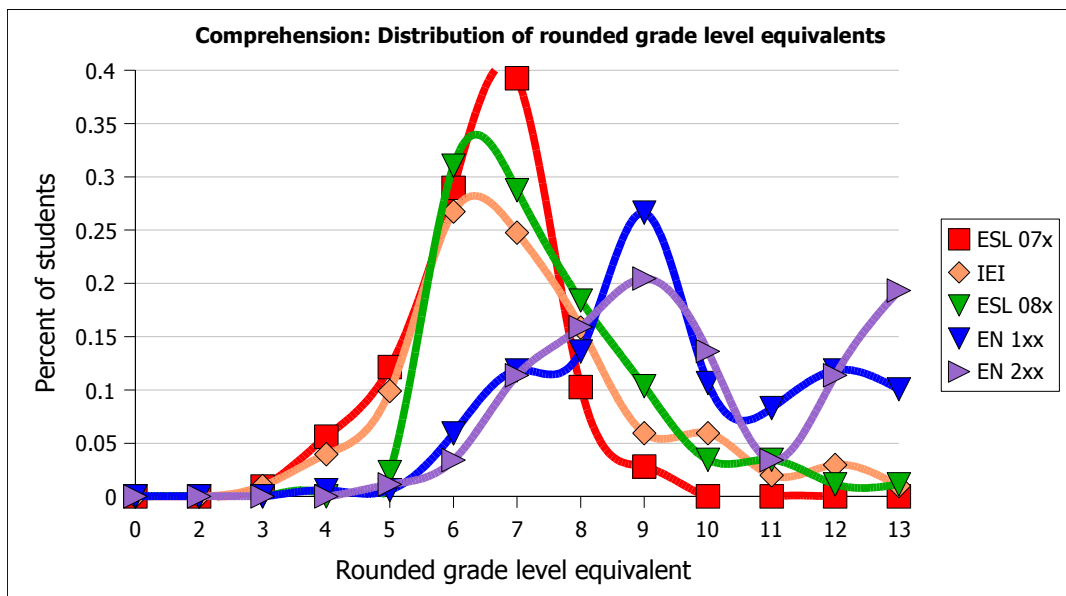
Course	Campus	Sample size n	Standard deviation	Mean (average)
ESL070	Chuuk	41	0.88	4.2
ESL071	Chuuk	66	0.91	4.53
ESL086	Chuuk	75	0.87	4.45
ESL088	Chuuk	13	0.73	4.16
ESL089	Chuuk	17	2.19	5.15
EN120b	Chuuk	11	1.49	5.72
SS100	Chuuk	14	0.88	4.21
FMI	FMI	31	1.05	5.17
UB	Kosrae	20	1.87	6.67
ESL086	Kosrae	4	1.77	5.35
ESL088	Kosrae	14	0.84	4.82
ESL089	Kosrae	17	1.75	5.62
EN110	Kosrae	11	1.12	5.94
EN120b	Kosrae	14	1.06	5.95
EN110	National	104	1.63	6.23
EN201	National	13	2.77	6.98
EN208	National	44	2.32	6.73
EN214	National	15	2.86	7.76
ESL088	Pohnpei	18	1.02	4.72
ESL089	Pohnpei	8	0.53	5.06
EN110	Pohnpei	11	1.15	5.94
ESL086	Yap	19	0.96	5.66
EN110	Yap	15	2.27	6.49
EN110/201	Yap	3	2.21	5.83
EN201	Yap	16	2.14	7.68

Table 2: Vocabulary grade-level equivalents



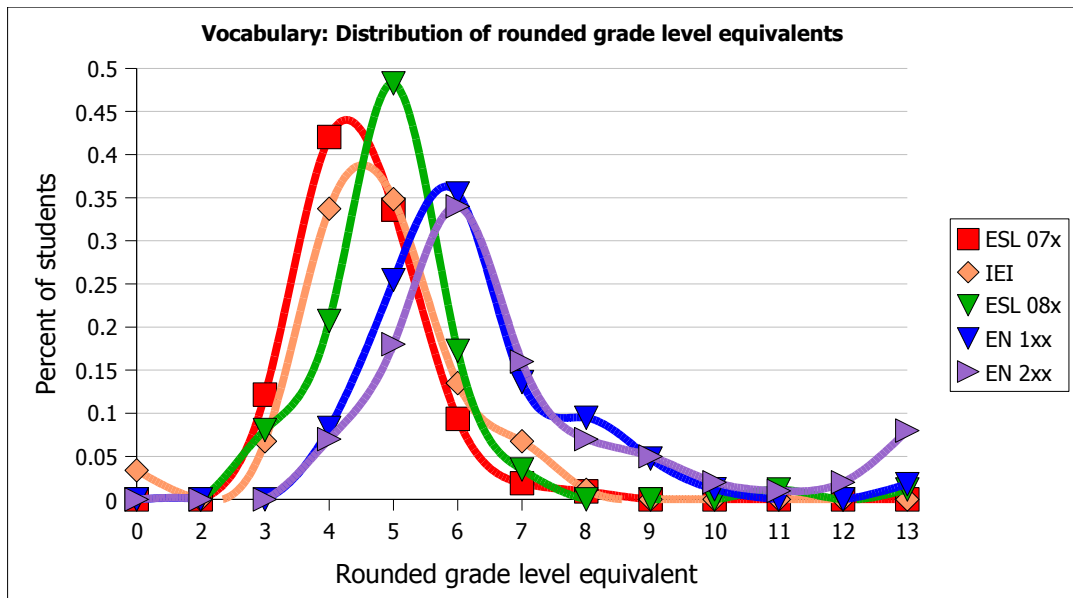
Bearing in mind that the Gates-McGinitie AR is not reliable below seventh grade, the large swath of mean scores below this level must be suspect at best, especially those below the fifth grade level. There is a risk that these means are random noise, not signals.

Comprehension: Distribution by course across all sites



ESL 07x, IEI, and ESL 08x share common peaks around seventh grade. The distribution has a single peak, somewhat normal. EN 1xx and EN 2xx are both bimodal with two peaks, one at ninth grade, one at 12th - 13th grade. All courses see a broad range of grade-level equivalents for their students.

Vocabulary: Distribution by course across all sites



All of the distributions are single peaked, and somewhat normal in distribution. All peaks are below seventh grade: Gates AR not designed to measure this range. Vocabulary is severely sub-adult at all levels. There is a statistical risk that results are near-random as the Gates AR is not designed to measure below the seventh grade level.

Comprehension and Vocabulary

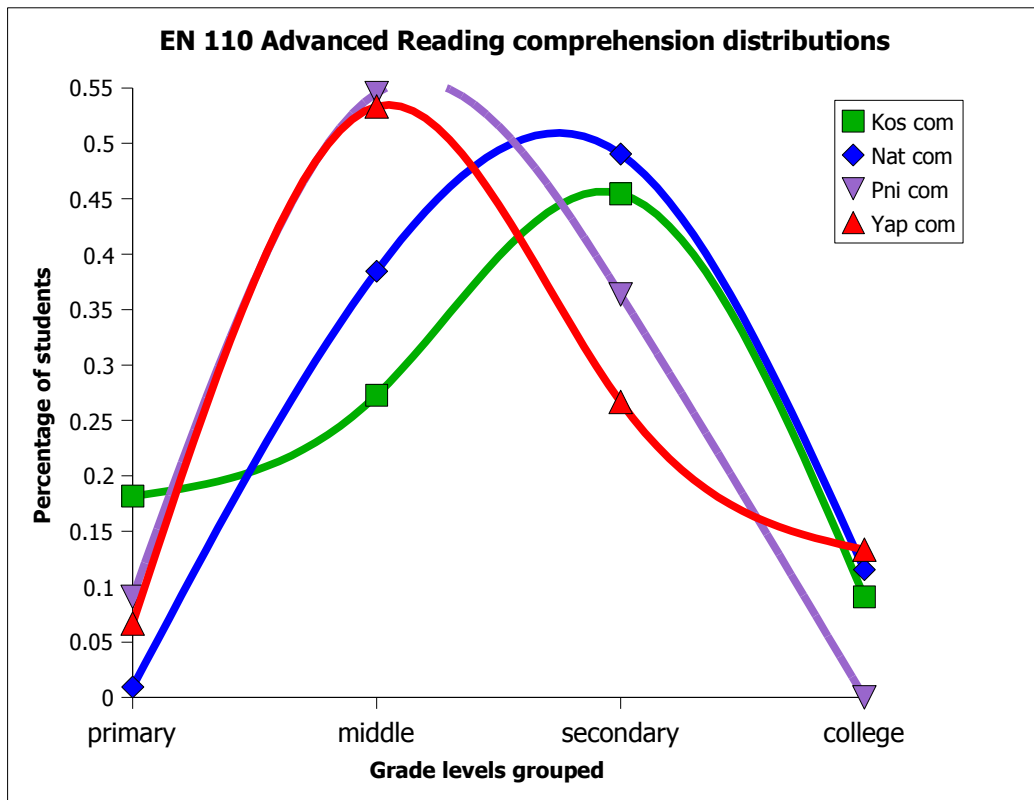
A standard candle is a reference to a light of a known brightness. Distance from that light can be gaged by the brightness of the light source. The dimmer the standard candle, the farther away the light.

The college lacks a course that acts as a standard candle per se, but EN 110 Advanced Reading would be a logical choice. This course is required of all students in all associate degree programs. The ability of the course to function at the same level on all sites virtually requires that the students in each classroom on each site have roughly equivalent comprehension and vocabulary skills.

At each grade-equivalent the sample sizes are too small to be significant or to generate a meaningful distribution. This author has chosen to regroup the grade levels by elementary (up to fifth grade), middle (sixth to eighth), secondary (ninth to twelfth), and college ("post-

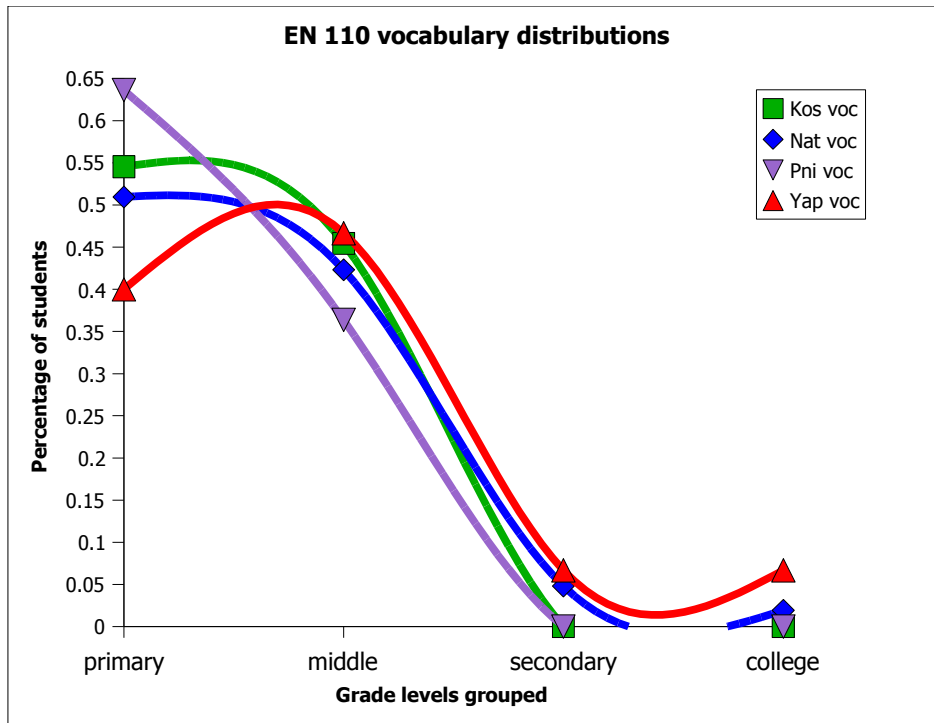
high school" on the actual Gates AR).

Distribution of the students by grade levels into the four groups above was done with an integer function: fractional grades were truncated to the integer. Thus a 5.9 is counted as elementary by the function and is not rounded up to middle school. This presumes that if 5.9 has any meaning it means a student at the end of fifth grade, not the start of sixth grade.



The National and Kosrae sites face similar distributions by grade-level equivalent peaking at secondary school level. Pohnpei and Yap sites face students with comprehension and vocabulary skills solidly in the middle school level. All sites have students who are functioning at an elementary school level in comprehension and vocabulary.

All sites except Yap face a majority of students in EN 110 whose vocabulary is at a primary school grade-level equivalent. No site has significant numbers of students at the secondary school level, let alone at a post-high school level of vocabulary.



Education is not a production facility, a factory. That said, a course where the "inputs" vary by site cannot be run in the same manner nor likely at the same "level" system-wide.

Of concern too from an "industrial age" perspective is the large range the distributions represent. A national site EN 110 instructor faces students with comprehension and reading grade equivalent levels that run from elementary to post-secondary school. Designing a course that challenges the most able students while meeting the needs of the weakest students is probably not possible. Designing a curricula that expects college level vocabulary is not possible.

Kosrae Upward Bound correlation substudy

The Gates AR grade-equivalents for the high school seniors in the Kosrae Upward Bound program afforded a rare rare opportunity to line up data and obtain correlations from a summer 2006 course in mathematics, the spring 2007 admissions examination results, and the Gates AR data. This enabled a small sample size study of correlations among all three measurements.

Correlation notes

A correlation of 1.00 is a perfect positive linear relationship. This is rarely seen in real data. A correlation above 0.70 is strong for this type of data.

A correlation between 0.60 and 0.69 could be considered moderately strong.

A correlation of 0.50 to 0.59 is moderate.

A correlation of 0.40 to 0.49 is weak but there is a relationship between the variables.

Below 0.4 the relationship is weak, with 0.00 being a purely random relationship. Nothing below 0.30 should be considered as anything other than noise.

Table row notes

- The table consists of correlation coefficients between various measurements on the different instruments.
- The first row are the variable names.
- The second row is each variable run against the Gates reading grade-level equivalent.
- The third row is each variable run against the Gates comprehension grade-level equivalent.
- The fourth row is each variable run against the Gates total grade-level equivalent.

Test	Grammar	Essay	95	96	99	100	Σm	COMET	u62	read	comp	total
read	0.43	0.36	0.28	0.04	0.3	0.31	0.34	0.45	0.49	1	0.5	0.87
comp	0.76	0.46	0.45	0.21	0.33	0.5	0.57	0.68	0.73	0.5	1	0.85
total	0.69	0.45	0.39	0.14	0.4	0.47	0.52	0.64	0.66	0.87	0.85	1

Table column notes

- Test is the Gates-AR subtest
- Grammar is the COMET grammar test variable.
- Essay is the COMET essay variable, 0 to 50 points spring 2007.
- 95, 96, 98, and 100 are the four math subtest scores.
- Σm is the sum of the math subtests - the total correct out of 40.
- COMET is the COMET score. Being a linear transformation of the z-score, the correlations match those of the z-score.
- u62 is a correlation run against the student's final percentage in the UB math course of summer 2006.
- The last three columns are the Gates AR variables themselves. Note the diagonal set of three ones - a variable correlated against itself yields perfect correlation.

When reading the following remember that correlation is not causation. The existence of a strong correlation does not mean one variable is the cause of the other.

Two results above indicate a strong correlation. The Gates comprehension grade-level equivalent showed strong correlation to the grammar section of the COMET. The Gates comprehension test also correlated strongly to the students percentage in the Upward Bound 2006 summer program mathematics course. The later is of interest in part because the course was a mathematics course, not a language or literature course.

This suggests that the grammar test does "behave" in a manner that correlates with the Gates

comprehension grade-level equivalent. Given the professional talent that works to produce the Gates test, this suggests that the college grammar section is "well-behaved." This is good news for the grammar subtest.

The essay test does not correlate well to the Gates measures, nor would one expect that a writing test would necessarily correlate well to a reading or comprehension test. Beyond needing to read the question, students can "end-run" vocabulary and comprehension limitations by using alternate wordings with which they are comfortable to communicate effectively.

That a relatively strong correlation was found between the Gates comprehension subtest and performance in a mathematics course a year earlier is an unexpected bonus.

Discussion

On a positive note, all sites show upward trends in comprehension and vocabulary as students move up through the sequence of courses. The college does have a positive impact on students comprehension and vocabulary.

The caveat to that positive note is that the students arrive at the college horribly underprepared for a collegiate curriculum. Comprehension levels are around eighth grade, vocabulary levels are around fifth grade equivalent. Texts in use at the college typically range from 10th grade level and up.

Of some interest in the ancient state site versus national site dialog is the position of the national site 100 and 200 level courses. These courses, along with EN 201 at Yap site, hold the highest means by rank order. This may be evidence that the national site attracts the highest academic achievers in the nation, although this is also explainable by the nature of admissions at the college. Half of the top 400 students admitted annually opt to attend the national site. Yet some of the other half do opt to attend their state site, leaving room for the suggestion that there is an academic synergy created at the national site.

The national site may be important as a gathering place for the academic achievers of the nation, arguably the future leaders of the nation in business, government, and academia. The national site remains a unique institution with students from four states. This environment may encourage English language acquisition and mastery.

A similar argument may actually hold for Yap site. Anecdotal observations of these author are that local caste issues often force Yapese students to resort to the neutral-tongue of English to communicate across caste barriers. The site, in bringing together students from across Yap, creates an environment that encourages the use of English.

Another group of students who attained a strong mean comprehension grade-level

equivalent were the Kosraen Upward Bound high school students. These students scored in the tenth grade-level equivalent for comprehension. While their vocabulary was in the sixth grade-level equivalent, this was actually a strong performance against the other subgroups. The Upward Bound program is a success by this narrow metric.

Vocabulary levels are clearly the greater weakness for the college students. With vocabulary levels somewhere in the middle of the elementary school deck, solutions that begin there come readily to mind. Reading programs that engage students and expand their working vocabulary, not just spelling lists, are clearly called for by this data. The college is not a cost effective place to lift reading levels from the fourth and fifth grade – or wherever they may lie below the seventh grade level.

Questions come to mind such as the vocabulary level of the teachers in those grades, and their reading capacity. Options such as a stronger library presence, programs to encourage parents and older siblings to read to children and younger siblings could also have a place in a comprehensive multi-year assault on the vocabulary levels.

A more haunting possibility is that the low vocabulary levels in the second language (L2) may be reflective in part of loss of the primary language (L1) among youth. Anecdotal evidence for this has been gathered in SC/SS 115 Ethnobotany and has been glimpsed in some of the entrance test essays where a student chooses to write in their L1 language and a reader familiar with the language has noted that the student is not using correct grammar, spelling, or vocabulary in their L1.

Closing comments

The hope is that this report will stimulate further questions, some of which might be answered by "slicing and dicing" the data in a different way, some of which might be answered by further research.

Tantalizing questions remain out of reach at this time, such as whether there are correlations between the Gates-McGinite AR scores and graduation rates, performance on the subtests of the COMET, retention, and grades. The Kosrae Upward Bound substudy hints at the data mining potential that is represented by all the assessment contact points that the college has with a student during their time in college programs. Upward Bound, for example, extends that data gathering capability to secondary school students prior to their taking the college entrance test.

A debt of thanks is owed to the many people who have worked long and hard on gathering this data. This work is a tribute to the capacity of the department of instructional affairs to carry out research and assessment.